| N A S A |

| N A S A |

USL / DBMS NASA / RECON

WORKING PAPER SERIES

Report Number

DBMS.NASA/RECON-14

The USL/DBMS NASA/RECON Working Paper Series contains a collection of reports representing results of activities being conducted by the Computer Science Department of the University of Southwestern Louisiana pursuant to the specifications of National Aeronautics and Space Administration Contract Number NASW-3846. The work on this contract is being performed jointly by the University of Southwestern Louisiana and Southern University.

For more information, contact:

Wayne D. Dominick

Editor
USL/DBMS NASA/RECON Working Paper Series
Computer Science Department
University of Southwestern Louisiana
P. O. Box 44330
Lafayette, Louisiana 70504
(318) 231-6308

(NASA-CR-184522) USL NASA/RECON PROJECT PRESENTATIONS AT THE 1985 ACM COMPUTER SCIENCE CONFERENCE: ABSTRACTS AND VISUALS Final Report, 1 (University of Southwestern Louisiana. Lafayette. Center for Advanced

N89-14962

G3/82 Unclas 0183560

4.15

1. Report No. IN-82	2. Government Accession No. 183560	3. Recipient's Catalog No.
4. Title and Subtitle USL/NGT-19-010-900: USL NASA/RECON PROJECT PRESENTATIONS AT THE 1985 ACM COMPUTER SCIENCE CONFERENCE: ABSTRACTS AND VISUALS		5. Report Date March 10, 1985
		6. Performing Organization Code
7. Author(s) FRANK Y. CHUM, SUZY GALLAGHER, MARTIN GRANIER, PHILIP HALL, DENNIS MOREAU, AND SPIROS TRIANTAFYLLOPOULOS		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Southwestern Louisiana The Center for Advanced Computer Studies P.O. Box 44330 Lafayette, LA 70504-4330		10. Work Unit No. 11. Contract or Grant No. NGT-19-010-900
12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered FINAL; 07/01/85 - 12/31/87
		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This Working Paper Series entry represents the abstracts and visuals associated with presentations delivered by six USL NASA/RECON research team members at the above named conference. The presentations highlight various aspects of NASA contract activities pursued by the participants as they relate to individual research projects. The titles of the six presentations are as follows:

(1) "The Specification and Design of a Distributed Workstation."
(2) "An Innovative, Multidisciplinary Educational Program in Interactive Information Storage and Retrieval,"
(3) "Critical Comparative Analysis of the Major Commercial IS&R Systems,"
(4) "Design Criteria for a PC-Based Common User Interface to Remote Information Systems,"
(5) "The Design of an Object-Oriented Graphics Interface," and
(6) "Knowledge-Based Information Retrieval: Techniques and Applications."

This report represents one of the 72 attachment reports to the University of Southwestern Louisiana's Final Report on NASA Grant NGT-19-010-900. Accordingly, appropriate care should be taken in using this report out of the context of the full Final Report.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Distributed Workstations, Educational Programs, Commercial IS&R Systems, User Interfaces, Object-Oriented Graphics Interface, Knowledge-Based Systems, Information Storage and Retrieval Systems	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	80	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

| N A S A |

USL NASA/RECON PROJECT PRESENTATIONS AT THE 1985 ACM COMPUTER SCIENCE CONFERENCE: ABSTRACTS AND VISUALS

Frank Y. Chum

Suzy Gallagher

Martin Granier

Philip Hall

Dennis Moreau

Spiros Triantafyllopoulos

Center for Advanced Computer Studies
University of Southwestern Louisiana
P. O. Box 44330
Lafayette, Louisiana 70504

March 10, 1985

| N A S A |

USL NASA/RECON PROJECT PRESENTATIONS

AT THE

1985 ACM COMPUTER SCIENCE CONFERENCE:

ABSTRACTS AND VISUALS

ABSTRACT

This Working Paper Series report comprises material presented by NASA/RECON research assistants at the 1985 ACM Computer Science Conference in New Orleans, Louisiana, March 12-14, 1985. Each entry includes both the abstract presented and the visuals used for the presentation.

The abstracts presented are:

(1) "The Specification and Design of a Distributed Workstation" by Frank Y. Chum,
(2) "An Innovative, Multidisciplinary Educational Program in Interactive Information Storage and Retrieval" by Suzy Gallagher,
(3) "Critical Comparative Analysis of the Major Commercial IS&R Systems" by Martin Granier,
(4) "Design Criteria for a PC-Based Common User Interface to Remote Information Systems" by Philip Hall,
(5) "The Design of an Object-Oriented Graphics Interface" by Dennis Moreau, and
(6) "Knowledge-Based Information Retrieval: Techniques and Applications" by Spiros Triantafyllopoulos.

TABLE OF CONTENTS

		Page
I.	THE SPECIFICATION AND DESIGN OF A DISTRIBUTED WORKSTATION	4
II.	AN INNOVATIVE, MULTIDISCIPLINARY EDUCATIONAL PROGRAM IN INTERACTIVE INFORMATION STORAGE AND RETRIEVAL	12
III.	CRITICAL COMPARATIVE ANALYSIS OF THE MAJOR COMMERCIAL IS&R SYSTEMS	26
IV.	DESIGN CRITERIA FOR A PC-BASED COMMON USER INTERFACE TO REMOTE INFORMATION SYSTEMS	48
V.	THE DESIGN OF AN OBJECT-ORIENTED GRAPHICS INTERFACE	58
VI.	KNOWLEDGE-BASED INFORMATION RETRIEVAL: TECHNIQUES AND APPLICATIONS	62

I. THE SPECIFICATION AND DESIGN OF A DISTRIBUTED WORKSTATION

by

Frank Y. Chum

The Specifications and Design of a Distributed Workstation. FRANK CHUM University of Southwestern Louisiana- This project includes the development of a general system for transparent sharing and access of resources in a distributed IS&R environment. The proposed PC-based distributed workstation (PCDWS) prototype will give IS&R users an integrated PC-based workstation environment for transparent access and sharing of resources available from both local and remote facilities. The PCDWS will provide a robust personal computer workstation environment with a comprehensive set of tools as functional components to serve the users, as well as intercommunication and uploading/downloading protocols between workstations and remote mainframes as well as between workstaticns, thus providing access to multiple local and/or remote DBMS and IS&R systems.

* This work was supported in part by the National Aeronautics and Space Administration under NASA Contract Number NASW-3846.

ACM Computing Reviews
Category Number(s)

B.4.3, C.2.4, H.4.3

Keywords

Distributed System
Communications
Information Search and Retrieval
Network Operations

DISTRIBUTED WORKSTATION SUPPORT

FOR

INFORMATION STORAGE & RETRIEVAL SYSTEMS

- GENERAL GOALS AND OBJECTIVES

*** Provide a mechanism for very wide distribution of the information storage and retrieval capabilities of NASA/RECON system.

*** Provide the potential performance improvement of performing selected functions local to the users.

*** Provide simulated information storage and retrieval system environments.

*** Provide state-of-the-art technology available to the NASA/RECON system.

- SPECIFIC R&D OBJECTIVES

*** Provide a robust personal computer workstation environment with a comprehensive set of tools as functional components to serve as a scientist's / engineer's R&D workbench.

*** Provide access to multiple DBMS and/or IS&R systems.

*** Provide distributed/networked workstation communication and uploading/downloading protocols between workstations and remote mainframes as well as between workstations.

RESEARCH & DEVELOPMENT METHODOLOGY

Phase I: Specifications

- User Requirement Analysis
- NASA/RECON Requirement Analysis
- Distributed Workstation Functional Specifications
- Evaluation of Candidate Workstation Systems
- Selection of Candidate Systems
- Model System and Network Architecture

Phase II: Design and Implementation

- Implementation Study and Design Specifications
- System Implementation
- Testing and Debugging
- Prototyping of Finished System

Phase III: Deployment

- Development Deployment and Support Strategies
- Operational Maintenance and Enhancement
- Performance Measurement and Evaluation

II. AN INNCVATIVE, MULTIDISCIPLINARY EDUCATIONAL PROGRAM IN INTERACTIVE INFORMATION STORAGE AND RETRIEVAL

by

Suzy Gallagher

An Innovative, Multidisciplinary Educational Program in Interactive Information Storage and Retrieval SUZY GALLAGHER University of Southwestern Louisiana-A description of the development of a set of transportable, college-level courses in the use of interactive, online IS&R systems, in particular the NASA/RECON system, is presented. The purpose of these courses is to educate science and engineering students in the effective use of automated scientific and technical IS&R systems. The presentation includes an overview of project objectives, management phases, and accomplishments to date. The methodology used for the course development is described and future plans, both long-term and short-term, are discussed.

* This work was supported in part by the National Aeronautics and Space Administration under NASA Contract Number NASW-3846.

ACM Computing Reviews
Category number(s)
(January 1984 revision)

H.3.0, K.3.2

List major key words
in order of importance

Education
Information systems education
Online information services

An Innovative, Multidisciplinary Educational Program in Interactive Information Storage and Retrieval

An Abstract of Thesis Research
Presented to
The Association for Computing Machinery
Thirteenth Annual Computer Science Conference

by
Suzy Gallagher
University of Southwestern Louisiana

March 12, 1985

PROJECT OBJECTIVES

Set of Courses

Hands-on Usage

Science and Engineering Students

Transportable Courses

MANAGEMENT PHASES OF NASA CONTRACT

A. Needs Analysis

B. Course Development

C. Pilot Course Administration

D. Pilot Evaluation

E. Development of Distribution Plan

F. Implementation of Distribution Plan

G. Conduct of Regional Seminars

H. Conduct of On-Site Seminars

I. Coordination of Request Processing and Information Dissemination

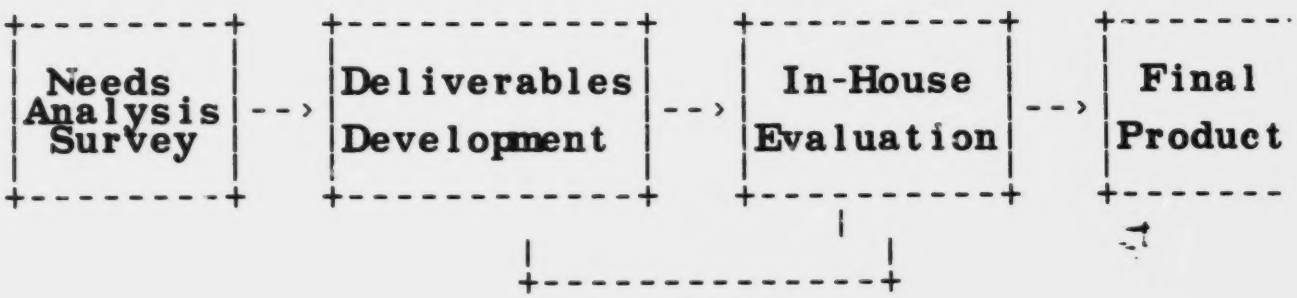
J. Course State-of-the-Art Enhancements

K. Institutional Surveys and Evaluations

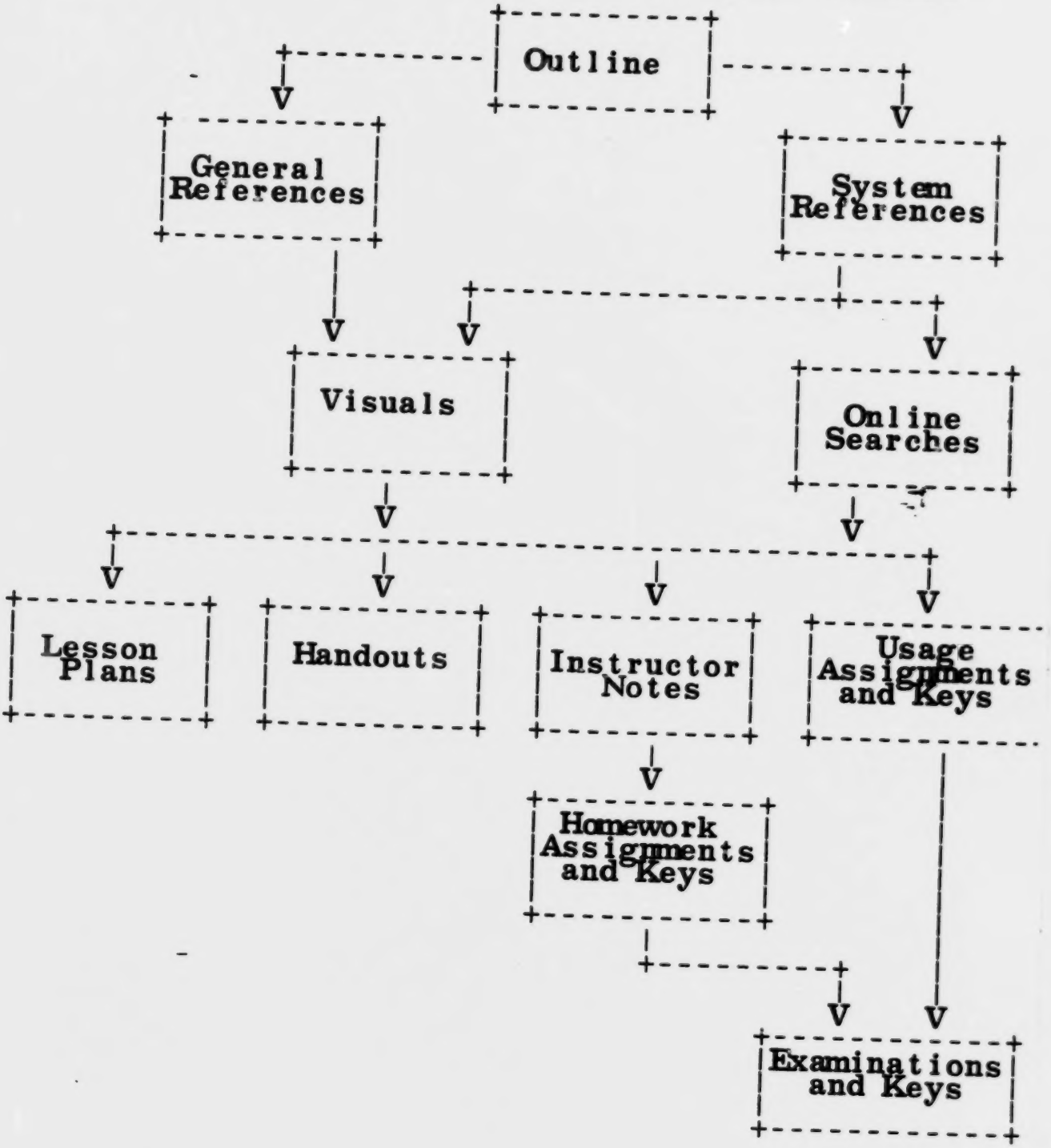
L. Graduated Student Surveys and Evaluations

M. Periodic Statistical Summary Reporting

OVERALL COURSE DEVELOPMENT DIAGRAM



COURSE DELIVERABLES DEVELOPMENT DIAGRAM



MAJOR CATEGORIES OF ACCOMPLISHMENTS

Project Control

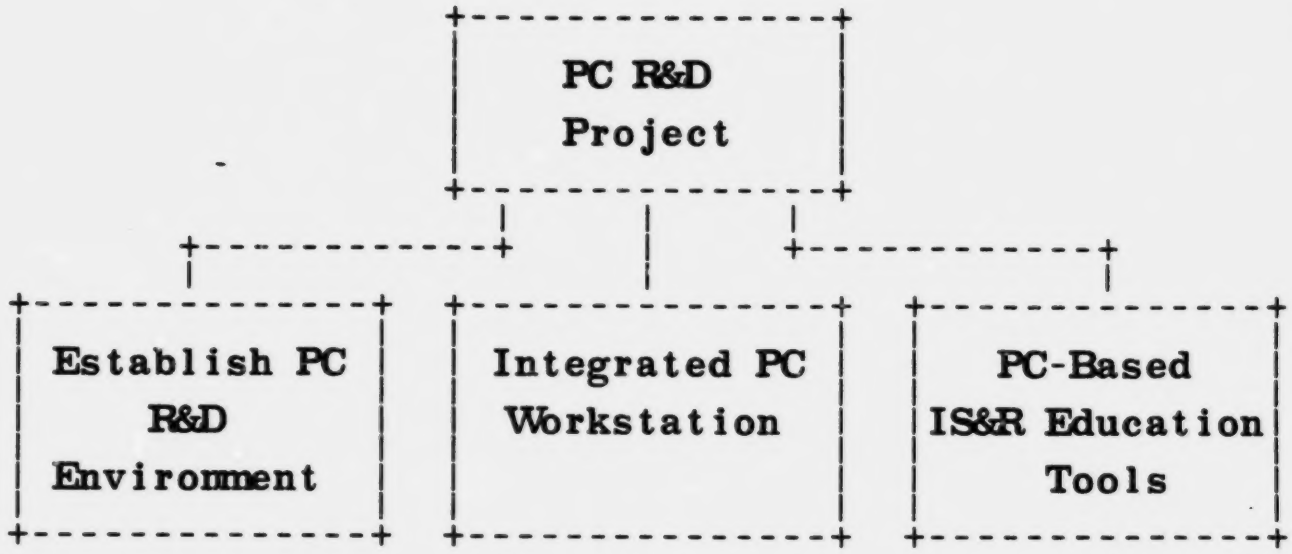
Needs Analysis

Project Working Paper Series

Course Development Working Paper Series

PC R&D Working Paper Series

Other Research Support



OBJECTIVES:

Continual Evaluation	Local Environment Interface	NASA/RECON Emulator
Develop Methods & Specifications	Remote Interface	IS&R Emulator Generator
Identify & Evaluate Projects	Distributed Interface	Interactive Presentation Development System
	Prototype PC Workstation	

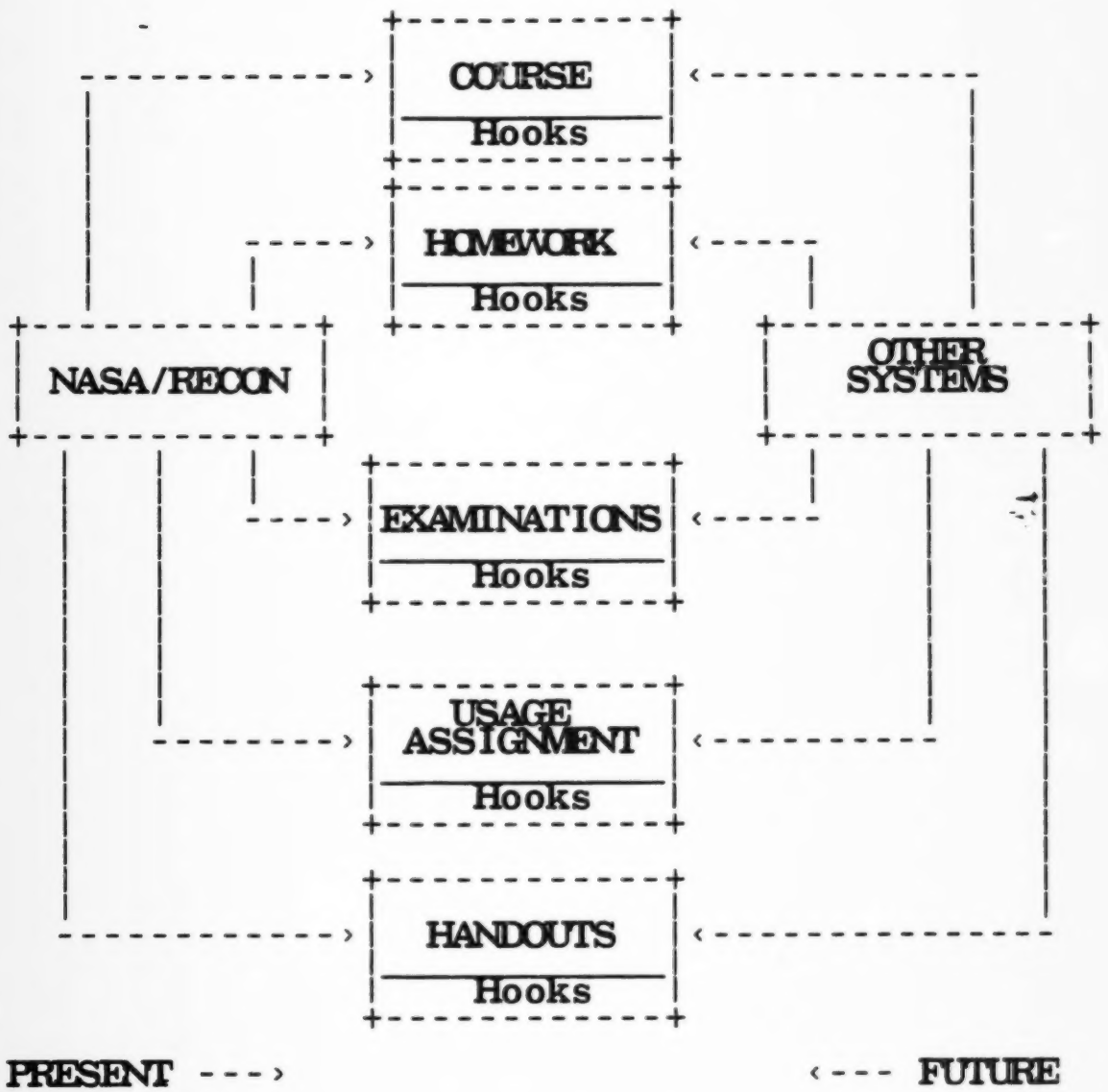
Relationship Between PC R&D Goals

FUTURE OF THE NASA/RECON EDUCATION PROJECT

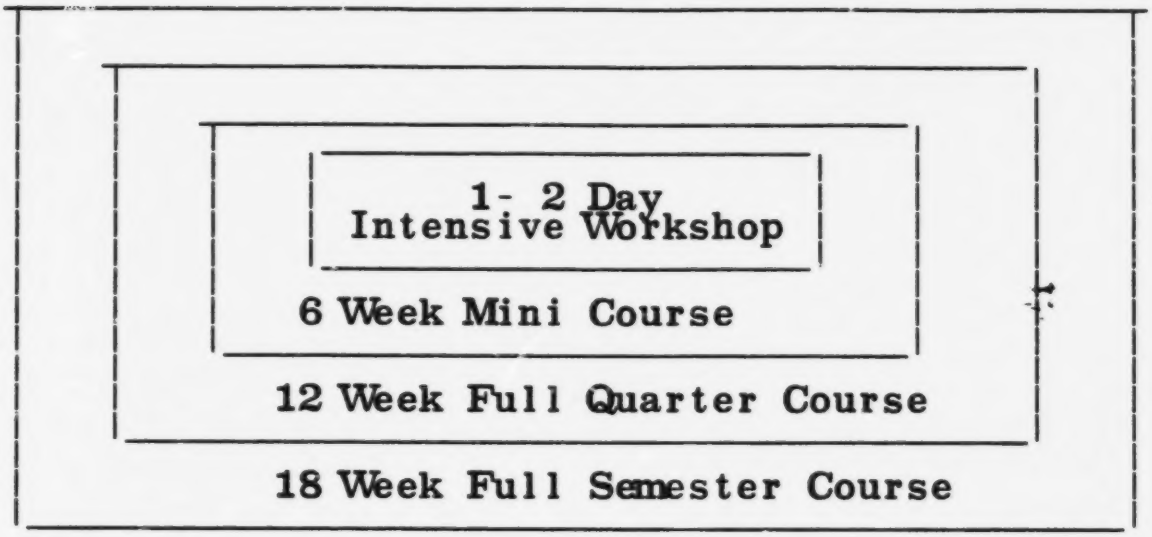
Short Term

- Define Additional Course Configurations
- Pilot Administration
- Pilot Evaluation
- Distribution Plan Development
- Additional Systems
- Additional Disciplines

SYSTEM INCORPORATION DIAGRAM



DEVELOPMENT PLAN FOR THE SET OF 4 COURSES



FUTURE OF THE NASA/RECON EDUCATION PROJECT

Long Term

- Distribution
- Evaluation
- Extensions
- Enhancements
- Non-Educational Institutions
- Request and Information Processing
- Result Reporting

CONCLUSIONS

End User Education

Co-ordinated Materials

Complete

Varied

Extensions

IS&R Systems

Disciplines

Enhancements

Improvements

Updates

Transportability

III. CRITICAL COMPARATIVE ANALYSIS OF THE MAJOR COMMERICAL IS&R SYSTEMS

by

Martin Granier

Critical Comparative Analysis of the Major Commercial IS&R Systems. MARTIN GRANIER University of Southwestern Louisiana- The number of online search services commercially available to the public today has, in many cases, complicated the search process for the casual user. Faced with an increasing number of systems, each one necessitating the mastery of a different retrieval language and a particular set of protocols, the user is left confused and often unable to satisfactorily complete his search. This study compares the syntax, semantics and functions of the different languages used by some of the commercial IS&R systems available in the United States today. It also leads to a proposal for a common command language which would cover the functions most needed during the retrieval process.

* This work was supported in part by the National Aeronautics and Space Administration under NASA Contract Number NASW-3846.

ACM Computing Reviews
Category number(s)
(January 1984 revision)
H.2.3, H.3.3, H.3.5, K.6.3

List major key words
in order of importance

Information Search and Retrieval
On-Line Information Services
Query languages
Software development

CRITICAL COMPARATIVE ANALYSIS

OF

THE MAJOR COMMERCIAL IS&R SYSTEMS

by

MARTIN GRANIER

Center for Advanced Computer Studies
University of Southwestern Louisiana
P.O. Box 44330
Lafayette, Louisiana 70504

PROBLEMS FACING

USERS OF INFORMATION SERVICES

- NUMEROUS AND VARIOUS SERVICES
  - SELECTING A SYSTEM
  - INTERROGATION LANGUAGES AND SEARCH STRATEGIES VARY
  - DIFFICULT TO REMEMBER MORE THAN A FEW COMMANDS
  - SAME DATABASES OFFERED ON DIFFERENT SYSTEMS
- SEARCH METHODS ON DIFFERENT SYSTEMS
  - LOGIN ACCESS
  - SEARCH QUERIES
  - DIFFERENT OUTPUT FORMAT

PROBLEMS FACING

USERS OF INFORMATION SERVICES (cont'd)

- STORING OUTPUT
- COMPARISON OF SEARCHES
- SUPPORT
- COMMUNICATIONS PROCEDURES
- CASUAL USERS
  - TIME CONSUMING
  - EXPENSIVE
  - FEAR OF USAGE

POSSIBLE SOLUTIONS

*** SEARCH SPECIALIST

*** USER INTERFACE

*** FRIENDLY GATEWAY

*** EDUCATING USERS

*** MULTIPLE LEVELS OF ACCESS

*** NATURAL LANGUAGE SEARCH

*** STANDARDIZATION: COMMON COMMAND LANGUAGE

THE ISSUES

- COMPATIBILITY
- LANGUAGE
- SAME INFORMATION ON DIFFERENT DATABASES

THE GOALS

- HELP THE CASUAL USER
- GIVE A BETTER AND MORE COMPLETE COVERAGE OF SEARCH
- COMPARISON OF COVERAGE, TIME AND PERFORMANCE
- ACCESS MANY DATABASES
- FACILITATE THE FORMATTING OF OUTPUTS

OUTLINE OF STUDY

- COLLECTION OF INFORMATION
- SELECTION OF SYSTEMS
- ANALYSIS OF AVAILABLE DATA
- PROPOSAL AND CRITICISM OF PROPOSED SET
- FIRST DRAFT OF COMMON COMMAND LANGUAGE
- APPLICABILITY TO EXISTING IS&R SYSTEMS
- POSSIBLE METHODS OF IMPLEMENTATION
- ANALYSIS OF RESULTS

COLLECTION OF INFORMATION

- SELECT LITERATURE
- SYSTEM USAGE
- EXCHANGE WITH USERS
- EXCHANGE WITH PROFESSIONAL SEARCHERS
- EXCHANGE WITH SYSTEM DESIGNERS
- USER MANUALS AND SYSTEM DOCUMENTATION

SELECTION OF SYSTEMS

- "MAJOR" ONLINE VENDORS
  - AVAILABLE TO A WIDE USER COMMUNITY
  - COMMERCIAL NETWORKS
  - SEARCH LANGUAGE IN ENGLISH AVAILABLE IN THE US
  - OFFERING A MINIMUM OF COMMERCIAL DATA BASES
- AVAILABLE PUBLICATIONS
- SELECTION OF COMMANDS

SYSTEM LIST

*** ERS

*** CAS-ONLINE

*** DIALOG

*** ISI

*** MEDLARS

*** ORBIT

*** PERGAMON-INFOLINE

*** QUESTEL

*** VU-TEXT

PROPOSAL AND CRITICISM OF

PROPOSED SET

*** ADVANTAGES OF CCL

*** DISADVANTAGES OF STANDARDS

*** DRAWBACKS OF CCL

ADVANTAGES OF CCL

*** DEGREE OF STANDARDIZATION BETWEEN SYSTEMS

*** WOULD BENEFIT:

USERS

PRODUCERS

PROCESSORS

*** TECHNOLOGICAL ADVANCES

- DISADVANTAGES OF STANDARDS

*** COST OF COMPLYING

*** RIGIDITY

*** SPECIAL CASES

*** INVESTMENT

*** DOUBLE SET OF COMMANDS

MUST BE ACCEPTABLE

*** NECESSARY COMPROMISE

*** CONFLICTS MAY EXIST

*** PREVIOUS ATTEMPTS (CCL, ISO)

FIRST DRAFT OF

COMMON COMMAND LANGUAGE

*** COMMON DEFINITION OF

RETRIEVAL SYSTEM FUNCTIONS

- COMMANDS TO INCLUDE (Cont'd)

*** DATABASE

*** DEFAULT ‹arguments›

*** DISPLAY

*** ERROR ‹message›

*** FIELD ‹labels›

*** LIMIT

*** LOGIN

*** LOGOUT

*** NEWS

*** REVIEW

*** SHOW

- DIFFERENCE BETWEEN LANGUAGES

*** COMMAND NAMES

*** COMMAND STRUCTURES

**** SEARCH MODE

**** COMMAND MODE

*** COMMAND FEATURES

*** COMMAND FORMATS

- COMPARATIVE STUDY

COMMAND LANGUAGES FOR VARIOUS IS & R

BRS	DIALOG	ORBIT	QUESTEL
_:	?	USER:	
root	expand	neighbor	display
and or not	and or not	and or not	and or not
adj with same and	(w) (s) (f) (c)	(w) adj (s) (f) (c) (r) and	adj sens nosens dst
print printoff	type print	print prt offline	type print pr
-	dialorder	order (supplier)	..order
off	logoff logoff hold	stop stop yes	logoff logoff hold

DIALOG	ORBIT	OTHERS
[logout/login]	[logout/login]	
* FILE file no.	FILE file name	(*) number
EXPAND term	NEIGHBOR term	
SELECT topic	FIND topic	SEARCH
SELECT AU = name	FIND name (AU)	SEARCH
SELECT a (C) b	FIND a AND b	SEARCH
COMBINE m * n	m AND n	
COMBINE m + n	m OR n	
TYPE x/2	PRINT	
TYPE x/6	PRINT TI	
PRINT x/2	OUTPUT OFFLINE	MAIL, PRINTOFF
TYPE n/2	Sn/OUTPUT	
?NEWS	NEWS	
LOGOFF	STOP (all done?)	DONE, QUIT LOGOUT, FINISH BYE, BYE-BYE

FUTURE RESEARCH

- APPLICABILITY TO EXISTING IS&R SYSTEMS
- POSSIBLE METHODS OF IMPLEMENTATION
- ANALYSIS OF RESULTS

CONCLUSION

WITHOUT STANDARDS

- NOT IN USERS INTEREST
- ANARCHY
- INHIBIT SEARCH
- LOSS OF CUSTOMERS
- LOSS OF REVENUES

WITH STANDARDS

- ALL QUERIES ASKED IN ONE WAY
- ABILITY TO SWITCH BETWEEN SYSTEMS
- PRESERVE PARTICULARITIES OF SYSTEMS
- NO CHANGE TO ORIGINAL DESIGN
- INCREASES ACCESS TO ONLINE SYSTEMS
- INCREASES USAGE OF ONLINE SYSTEMS

IV. DESIGN CRITERIA FOR A PC-BASED COMMON USER INTERFACE TO REMOTE INFORMATION SYSTEMS

by

Philip Hall

Design Criteria for a PC-Based Common User Interface to Remote Information Systems PHILIP HALL University of Southwestern Louisiana- The problems associated with retrieval by casual users of information stored in remote IS&R systems and the possible utilization of personal computers to solve these problems are discussed. A standardized system which will allow the user to access information stored in many distinct systems through a single common interface is described. The intent of this system is to spare the user the necessity of learning multiple command languages in order to access multiple systems and also retain the full retrieval capabilities of each system. Several levels of interaction are provided to facilitate new user learning phase activity and allow the intermediate and advanced users to interact with the system with the minimum necessary prompting. The system is designed to maximize utilization of local processing and display capabilities and to provide built-in evaluation tools.

* This work was supported in part by the National Aeronautics and Space Administration under NASA Contract Number NASW-3846.

ACM Computing Reviews
Category number(s)
(January 1984 revision)

H.1.2, H.3.3, H.3.5, K.8

Major Key Words:

User Interface Design
User/System Interaction
Information System Interfaces
Personal Computers
Online Information Services

USL NASA
Personal Computer
Research
and
Development

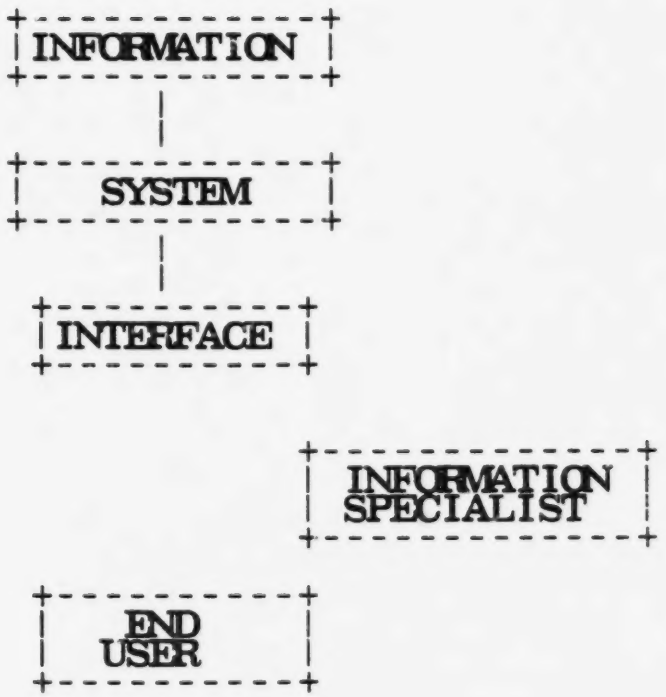
PC-BASED COMMON INTERFACE TO MULTIPLE REMOTE IS&R SYSTEMS

DEFINING THE CASUAL USER

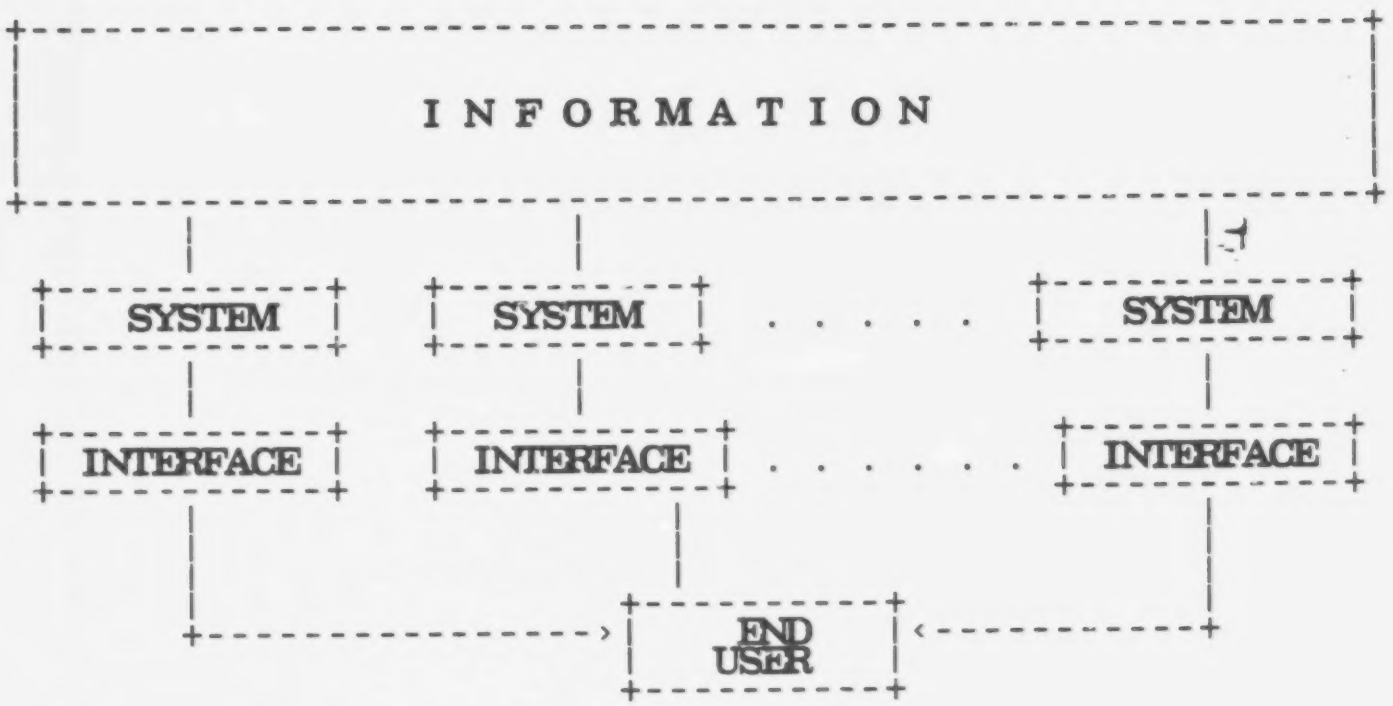
- 70% of User Population
- Characteristics
  - No Desire to Memorize Command Languages
  - Infrequent Access to System
  - Limited Knowledge of Programming
  - Limited Knowledge of Command Languages
  - Extensive Knowledge of Subject Field
  - IS&R Access not REQUIRED By Job
  - Job Enhancement Thru IS&R Access

TRADITIONAL

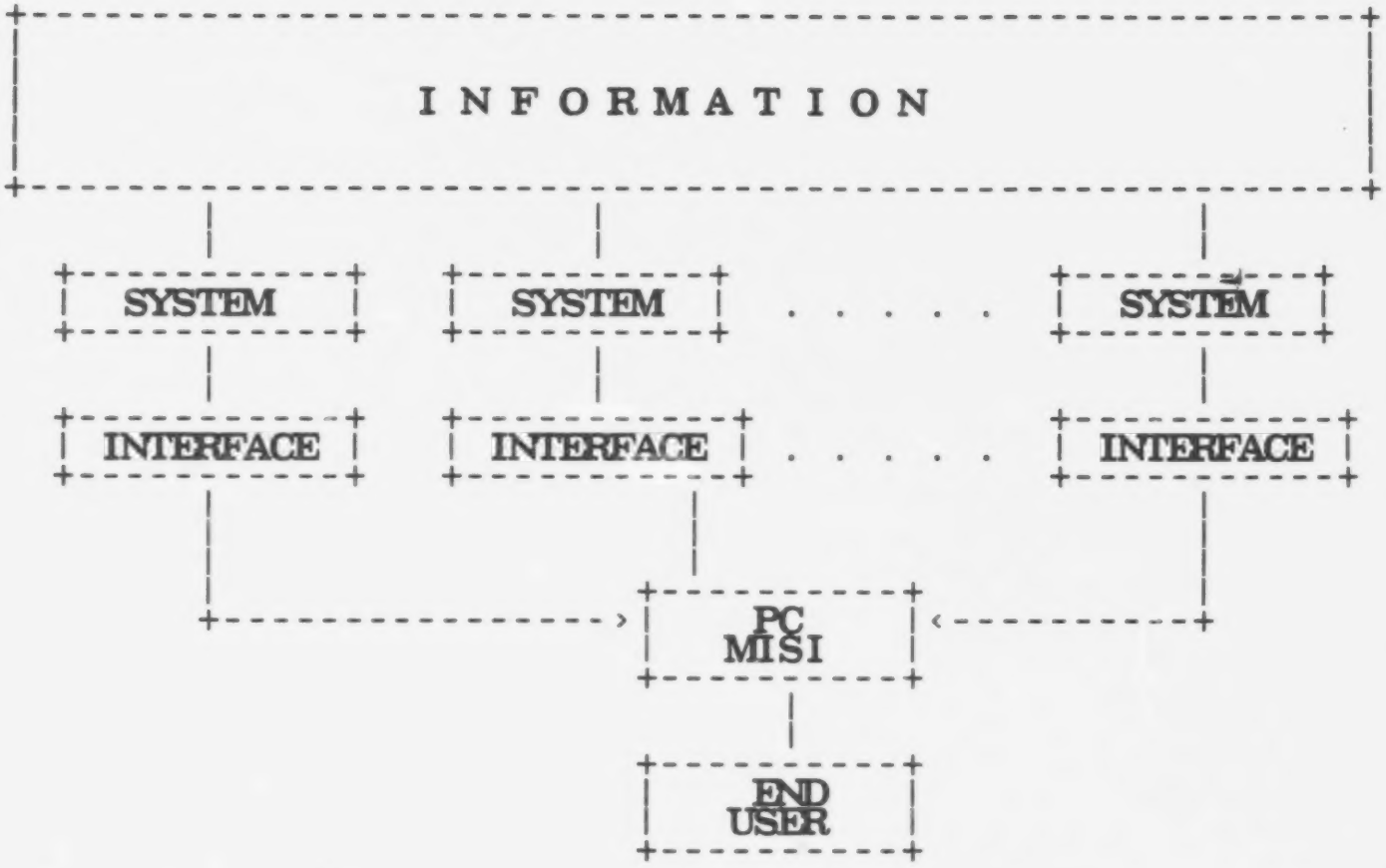
INFORMATION SPECIALIST USAGE



END-USER MULTIPLE SYSTEM USAGE



COMMON INTERFACE



DESIGN GOALS

- Ease of Access to Different Systems
  - Handle Communications
  - Facilitate Addition of New Systems
  - Incorporate Changes to Existing Systems
- Improve User/System Interaction
  - One Interface to All Systems
  - System Documentation
  - Multilevel Design
  - Facilitate Downloading

FUTURE ENHANCEMENTS

*** Expert System / AI Applications

*** Enhance User Queries

*** "Advice" to Users

*** "Find" System With Information

V. THE DESIGN OF AN OBJECT-ORIENTED GRAPHICS INTERFACE

by

Dennis Moreau

-The Design of an Object Oriented Graphics Interface. DENNIS MOREAU University of Southwestern Louisiana- A description of an advanced graphics interface design that provides the applications developer with a very high level graphics environment is presented. The object oriented design is shown to be appropriate to achieving device and implementation independence. This approach is also shown to provide a flexible means of managing non-graphic information associated with graphic objects. Implementation, using standard graphics primitives, is proposed.

* This work was supported in part by the National Aeronautics and Space Administration under NASA Contract Number NASW-3846.

ACM Computing Reviews
Category Number(s)
(January 1984 revision)

H.1.2, I.3.2

List major keywords
in order of importance

Graphics Interface
Object oriented environment

An Object-oriented Graphics Interface

Objectives:

To provide an easy to use interface for graphics applications development.

To provide conceptually consistent structure for partitioning graphics functionality.

To provide an extensible environment for archiving graphics components and tailoring available functions.

Interface Evaluation

Test Applications

Interactive Presentation Development System

3-D Surface Construction and Display System

Dataflow Programming Workstation

Evaluation Parameters

Development Time

Application Port Complexity

Performance Degradation

VI. KNOWLEDGE-BASED INFORMATION RETRIEVAL: TECHNIQUES AND APPLICATIONS

by

Spiros Triantafyllopoulos

Knowledge-Based Information Retrieval: Techniques and Applications - SPIROS TRIANTAFYLLOPOULOS University of Southwestern Louisiana - A collection of knowledge-based tools and techniques in support of information retrieval is presented. Applications such as natural language query systems, data integrity and concistency control, and intelligent interfaces to remote systems and DBMS front-ends, are discussed. Knowledge-based tools and techniques in support of such applications, including frame- and rule-based knowledge representation, knowledge acquisition and utilization, and intelligent dictionaries are presented. The presentation includes future research issues and extensions to existing applications and methods.

* This work was partially supported by the National Aeronautics and Space Administration, under NASA Contract Number NASW-3846.

ACM Computing Reviews
Category Number(s)
(January 1984 revision)

H.2.3, H.3.3, I.2.1

Key words:

Knowledge-Based Systems
Information Systems
Software Tools and Techniques

Knowledge-Based Information Retrieval: Techniques and Applications

by

Spiros Triantafyllopoulos

Center for Advanced Computer Studies

University of Southwestern Louisiana

Lafayette, LA. 70504

THE RETRIEVAL PROCESS

- Task: Data or Information Retrieval
- Available Solutions:
  - Application Programs
  - "Canned" Command Files
  - Menu Systems
  - Programming Languages
  - Query Languages
- Oriented towards the Frequent User
- Casual User has to Invest:
  - Time
  - Effort
- Typically, too many Rules, Languages and Terms

THE RETRIEVAL PROCESS (Cont'd)

- Specialized Software System Knowledge Required from the User's part
- The Software System itself does not have any Knowledge of the Human's Process of Thought
- The Result: Knowledge Gap between User and System
- The Solution: Make a Part of the User's Knowledge available to the System

KNOWLEDGE ISSUES

*** Knowledge Representation

    *** The encoding of domain-specific knowledge in a machine-readable and usable format

*** Knowledge Utilization

    *** The use of knowledge by the DBMS to assist the user in the retrieval process

*** Knowledge Acquisition

    *** The acquisition of knowledge at either

        --- Program development time

        --- Program initialization time

        --- Program execution time

KNOWLEDGE REPRESENTATION

- Frame-Based
  All Knowledge about an entity in one "slot"

- Rule-Based
  Allowable Constructs specified by rules

- Knowledge Storage and Retrieval
  - Typically, a Separate KBMS (Knowledge Base Management System) is used
  - This approach uses the Host DBMS/FMS for Knowledge Storage and Retrieval
  - Knowledge transformations needed
    - Frame-Based knowledge made to fit into table form (relationships)
      - --- Normalization
      - --- Redundancy Elimination
      - --- Storage Efficiency
    - Rule-Based knowledge also transformed to "fit" in table form

KNOWLEDGE REPRESENTATION (Cont'd)

... An Example of Knowledge Representation

NAME: salary
TYPE: noun
DATATYPE: numeric, real
RANGE: 0.00 to 99,999.99
PATTERN: { $ } [0-9]*{ .[0-9]* }
USED BY: raise, pay, add, subtract
IS-A: object
APPLICABLE-TO: employee
OPERATORS: ge, le, gt, lt, eq, ne
UNIT: $, dollar, dollars
SYNONYMS: payment, paycheck, pay

... Can be represented in relations as follows

KNOWLEDGE REPRESENTATION (Cont'd)

****** Proposed Knowledge Representation *******

noun frame

name	type	datatype	max	min	pattern	unit

synonyms representation

term	stands for

verbs representation

verb	subject	object

adjective representation

adjective	noun	implies

dictionary representation

word	word type

multiword representation

term	pattern id	rank

KNOWLEDGE REPRESENTATION (Cont'd)

*** In addition, the following knowledge is needed to process the user requests

*** Allowable Syntactical Constructs

--- Since it is fairly constant, syntax is "hardcoded" in pattern form

--- Verification of input pattern via matching with allowable patterns

*** Allowable Lexical Terms

--- Also stable, "hardcoded" in the program

--- Rules that transform different grammar forms in the ones known by the system (i.e., different tenses, singular/plural forms, etc.)

KNOWLEDGE UTILIZATION

- Knowledge is abstracted as
  - nouns
  - verbs
  - adjectives
    - sequences of words (idioms, etc.)
    - noicewords
  - overall organization (DB schema)
- With such abstraction, it is relatively simple to design and implement the knowledge processor for the restricted natural language
- Abstraction hides all details of storage/retrieval
- Knowledge processor asks for knowledge from the DBMS-implemented KBMS

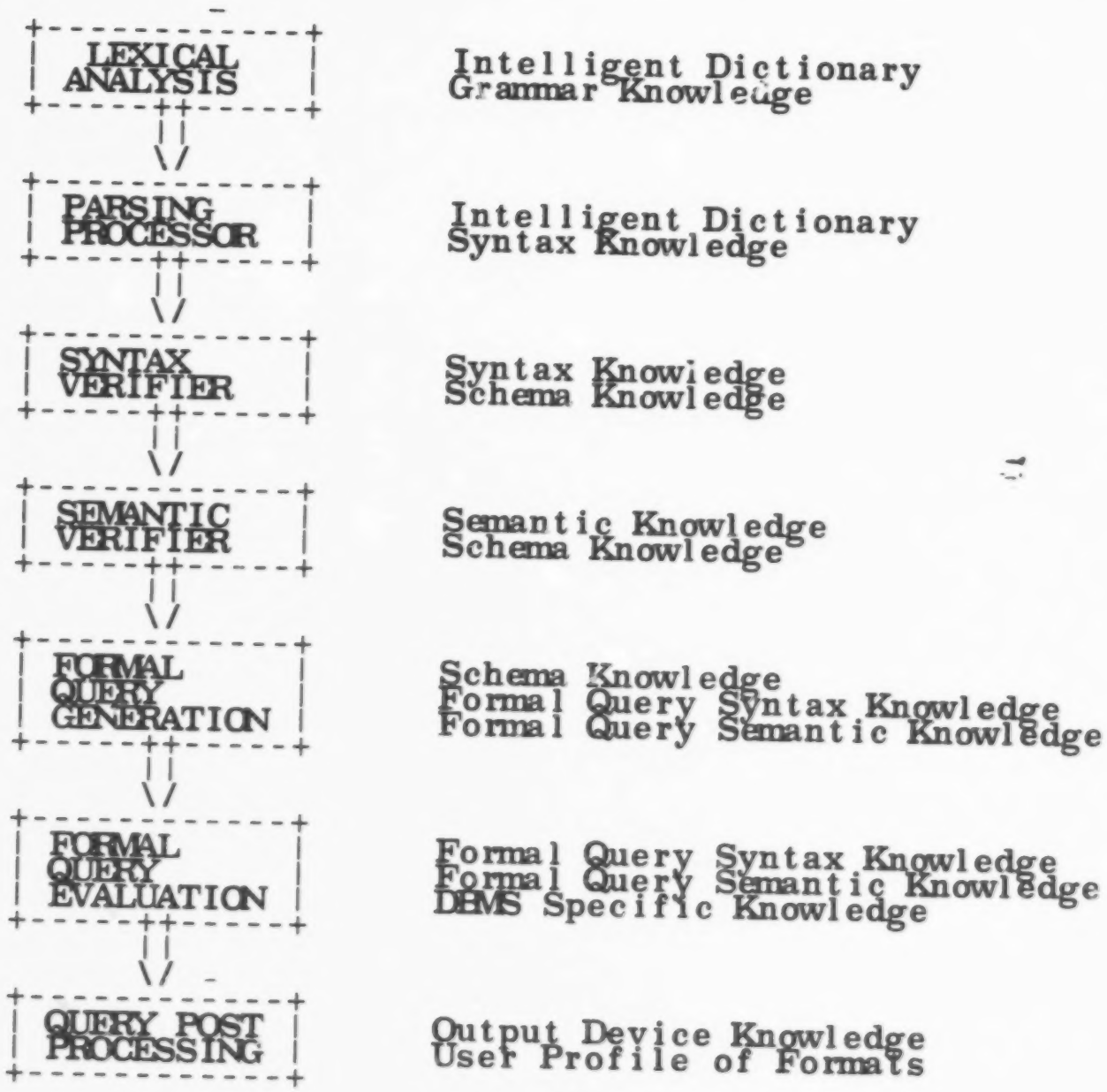
KNOWLEDGE UTILIZATION (Cont'd)

- Natural Language Processing
  - Utilizing Domain Vocabulary knowledge and "natural" language constructs
    - --- Lexical Analysis
    - --- NL Query Parsing
    - --- Syntactic Verification
    - --- Semantic Verification
    - --- Formal Query Generation
- Formal Query Evaluation
  - Handling of Null Results
  - Query Simplification
  - Query Optimization
- Other Usage
  - Format Selection
  - Full Report Generation

KNOWLEDGE UTILIZATION (Cont'd)

- Knowledge Abstraction Facilitates General System Design
  - Modularity
  - Granularity
  - Portability
  - Variety of Applications
- Knowledge Utilization Modules
  - Require no special languages i.e., Prolog, Lisp, etc.
  - Require no special Environment
  - Require no special knowledge in the implementation phase
  - Use simple algorithms

KNOWLEDGE UTILIZATION (Cont'd)



KNOWLEDGE ACQUISITION

- Knowledge is acquired at
  - System Design Time
    - Grammar Rules
    - Grammar Transformation Rules
    - Syntactic Patterns
    - Syntactic Rules
  - System Initialization Time
    - Schema-based Knowledge
    - Domain-specific Knowledge
    - Dictionary Initialization
      - Basic Knowledge needed for the system to become operational
      - Common Knowledge (applicable to a group of domains)
    - Transformation of Schema

KNOWLEDGE ACQUISITION (Cont'd)

- At System Run Time
  - --- System can "learn" new entries
    - --- If a term is unknown, after all transformations, system asks
    - --- If application domain is expanded system can learn the new terms
- At System Maintenance Time
  - --- Updates to present Knowledge
  - --- Corrections to present Knowledge
  - --- New definitions/terms

KNOWLEDGE ACQUISITION (Cont'd)

- Access problems have to be solved
  - Who has access to add knowledge
  - Who has access to modify knowledge
- Who is responsible for initializing the KBS
- How the system can be tailored for use at a Production environment
- How the system can be maintained at such an environment

CONCLUSIONS

*** Advantages

*** Increased applicability

--- Knowledge about the application is not hardcoded in the system, but can be expanded/tailored

*** Increased system flexibility

--- No specific KBMS Needed

--- No specific DBMS Needed (Minimal code changes in the formal query generation module)

--- No specific environment

--- Modularization and abstraction

CONCLUSIONS (Cont'd)

- Simpler system
  - --- Design time is reduced
  - --- Implementation time is reduced
  - --- No specific implementation
  - Knowledge needed
- Disadvantages
  - Storage efficiency
    - More required due to normalization/ transformations in relational form
  - Execution efficiency
    - Slower due to database accesses for knowledge retrieval